

（深圳）

实验报告

开课学期： 2021秋季

课程名称： 大数据导论

实验名称：Hadoop环境配置与基本操作

实验性质： 设计型

实验学时： 2 地点： T2608

学生班级： 1901105

学生学号： 190110509

学生姓名： 王铭

评阅教师：

报告成绩：

实验与创新实践教育中心制

2021年11月

# 实验目的

1. 熟悉Hadoop分布式配置的方法。

2. 熟悉命令行运行Mapreduce作业的原理和操作。

# 实验内容

1. 根据实验指导书，小组合作搭建Hadoop分布式环境并截取WebUI界面。

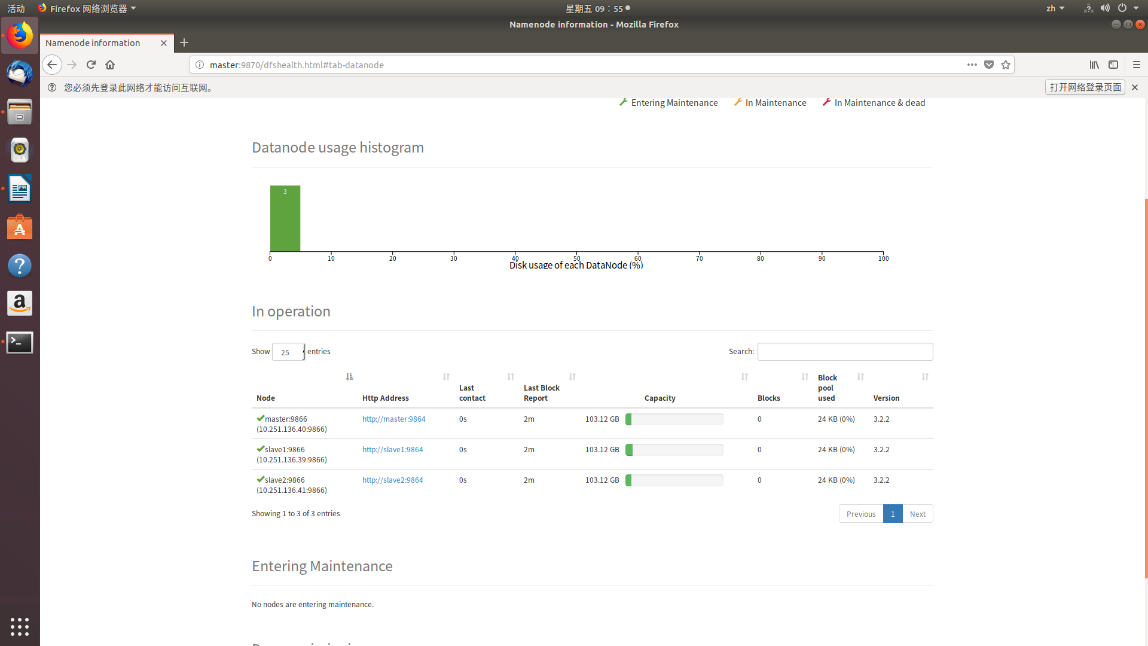
2. 配置Hadoop客户端，实现文本文件的上传。

3. 为WordCount项目添加集群运行配置，运行并截图。

# 实验过程

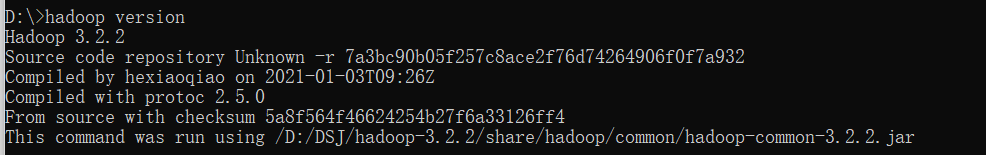
1.hadoop分布式环境搭建

小组成员分为master，slave1，slave2进行操作，我负责master节点配置。按照实验指导书详细内容顺序配置，完成搭建。WEB截图如下：



2.配置Hadoop客户端，实现文本文件上传

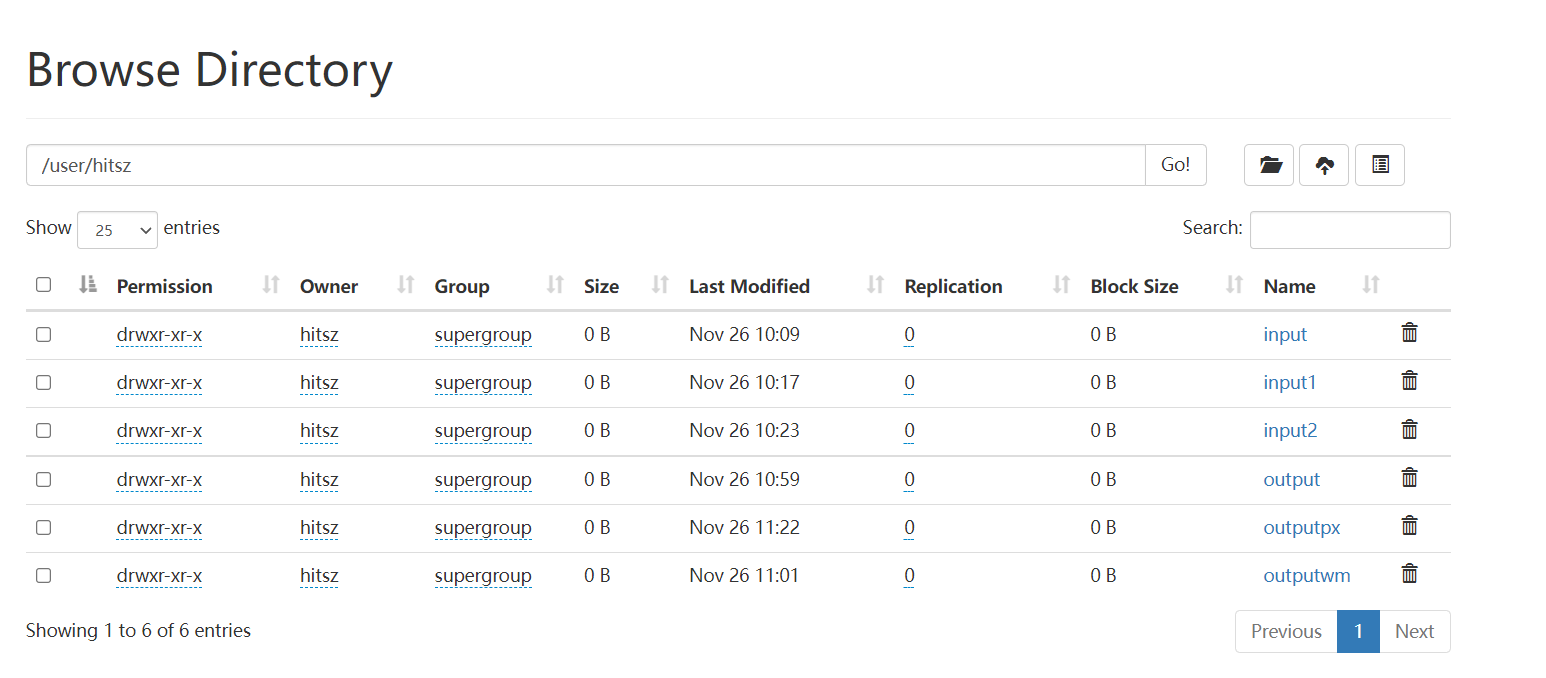
配置Hadoop客户端在作业二中已经实现。截图如下：

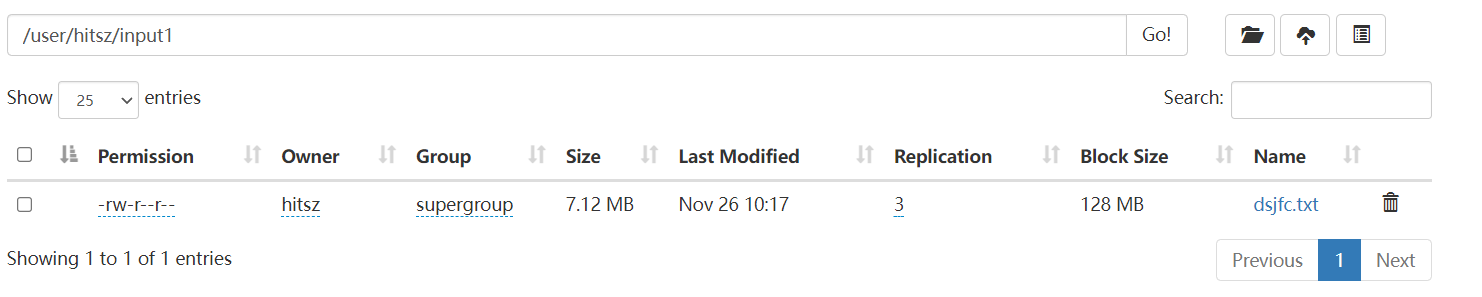


搭建好hadoop分布式环境后，在客户端的命令行中输入上传命令：

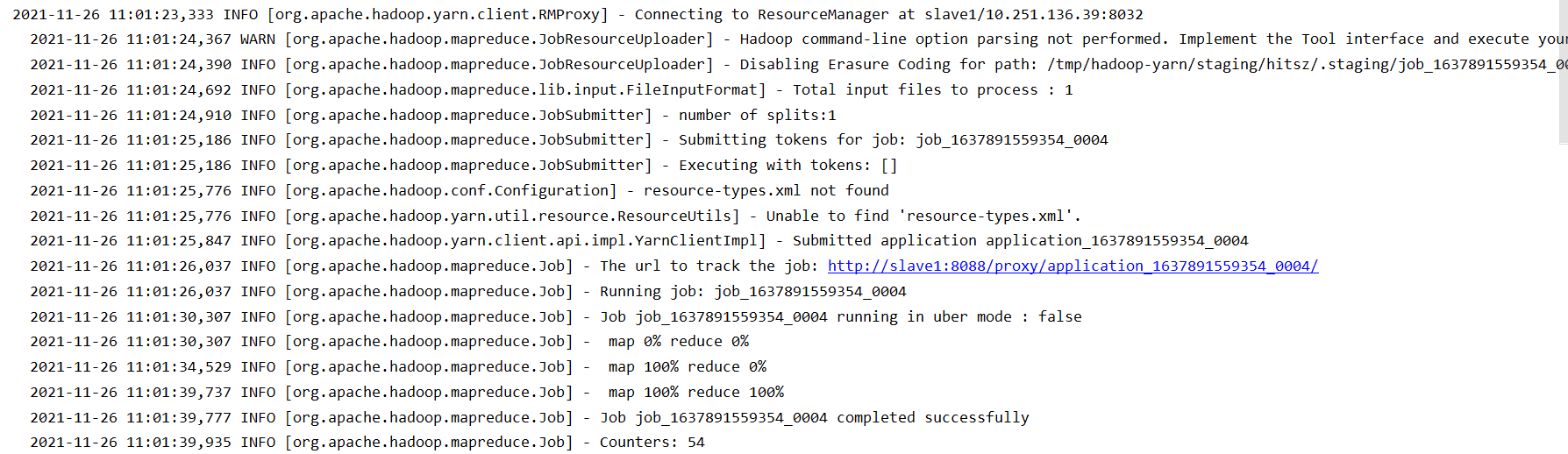


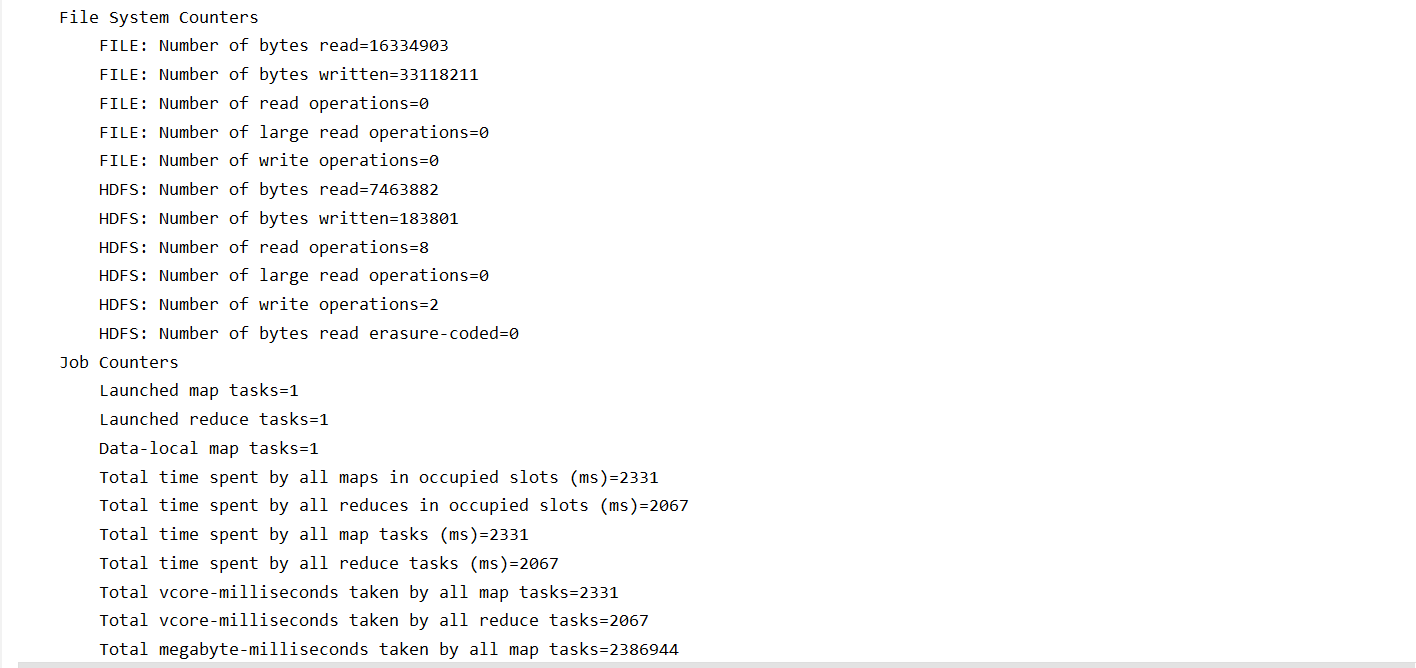
查看web端结果，显示有input1，上传成功，截图如下：

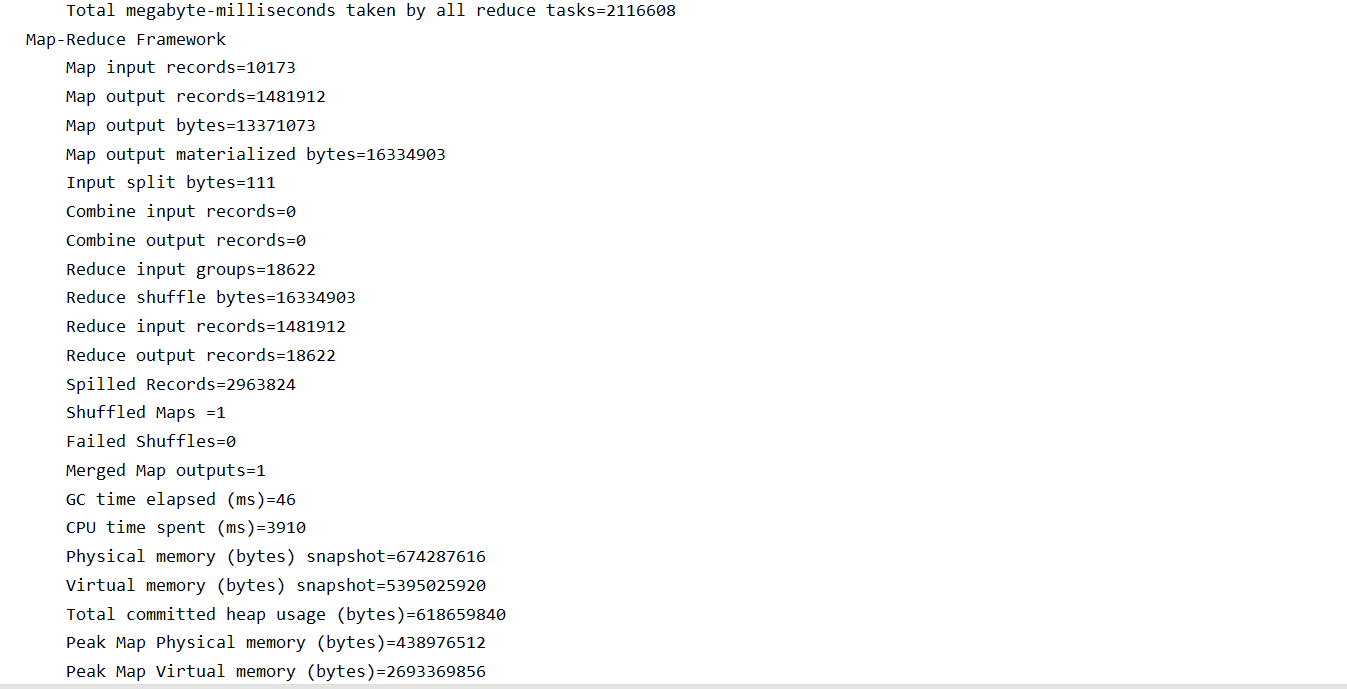


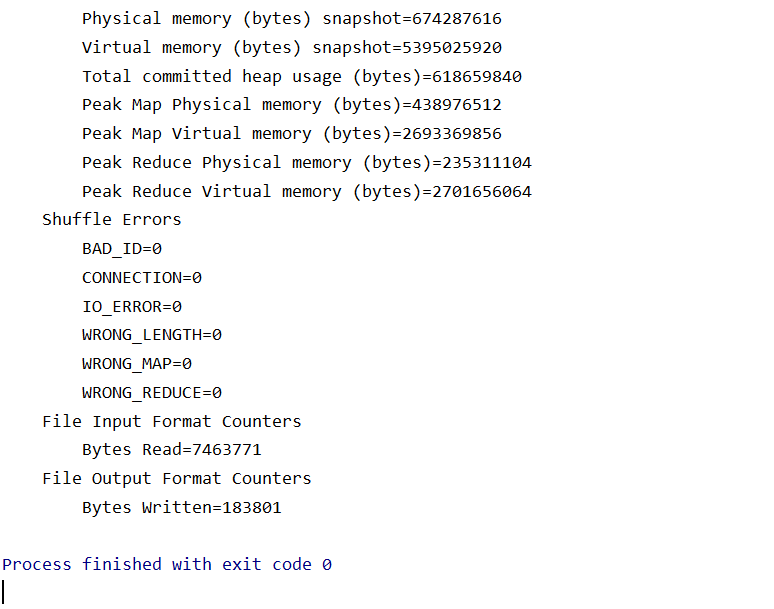


执行WordCountDriver程序，截图如下：

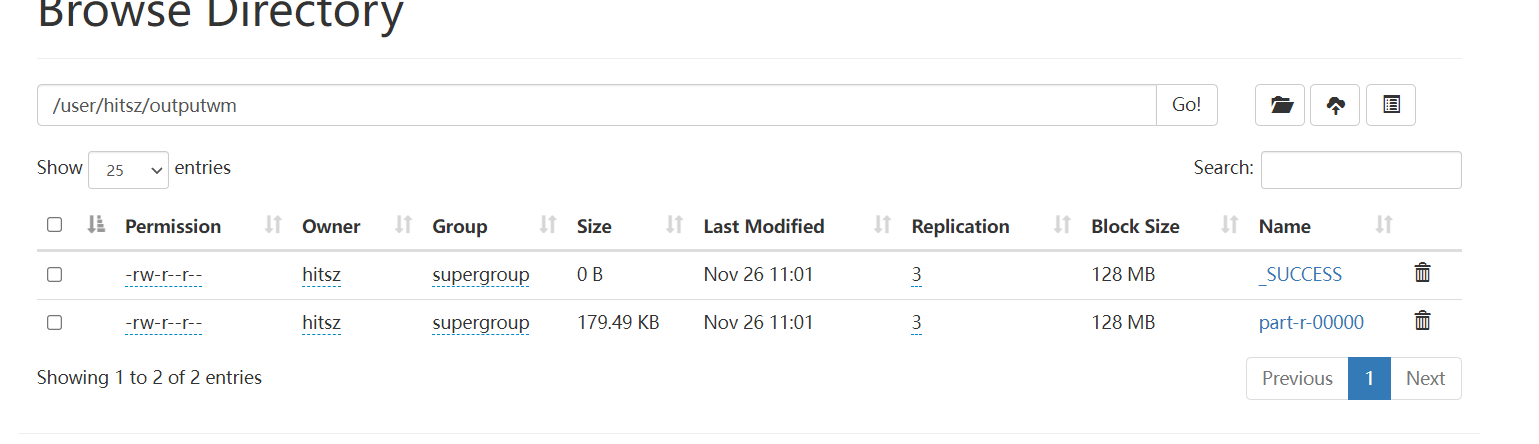








查看web端结果：

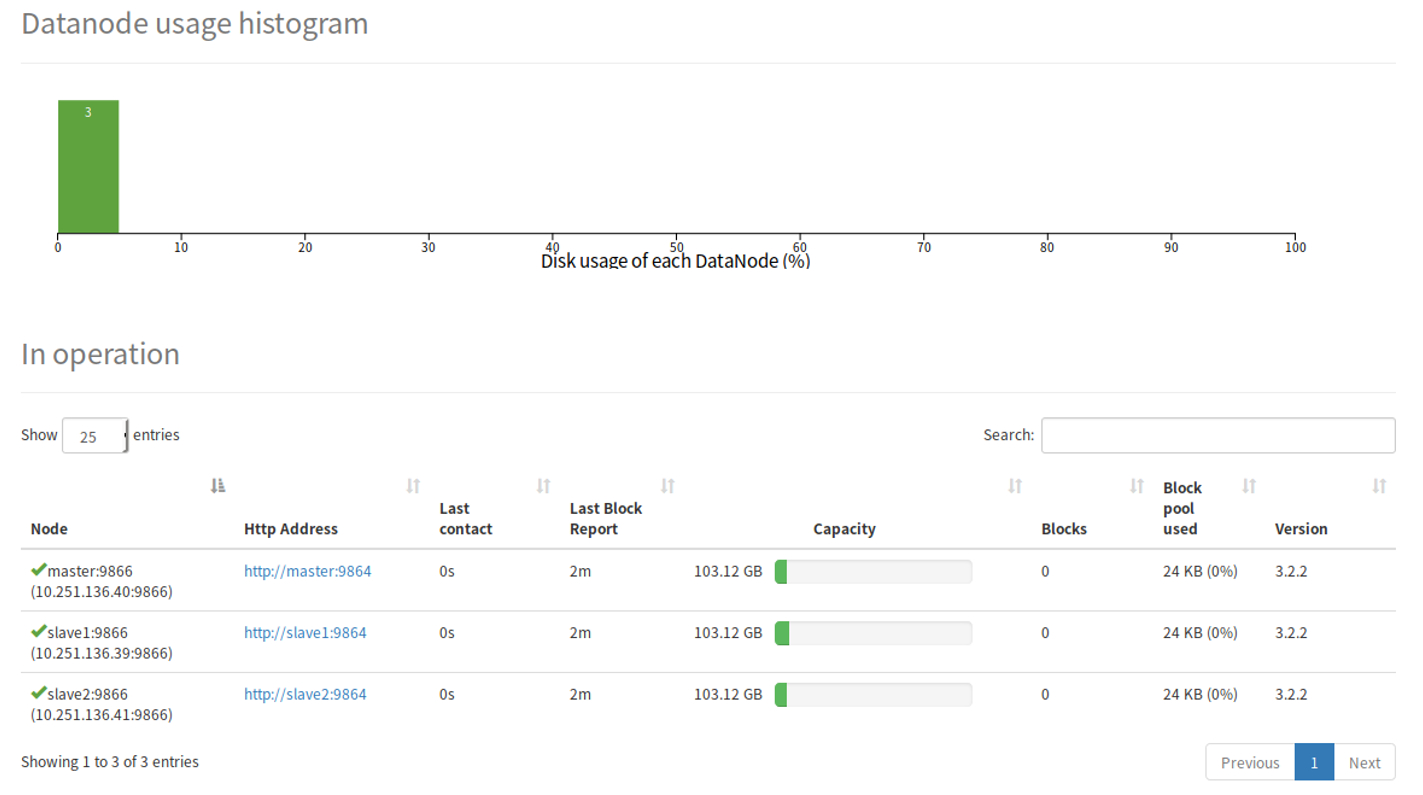


下载part-r-00000文件，查看mapreduce的操作结果，其命令如下：

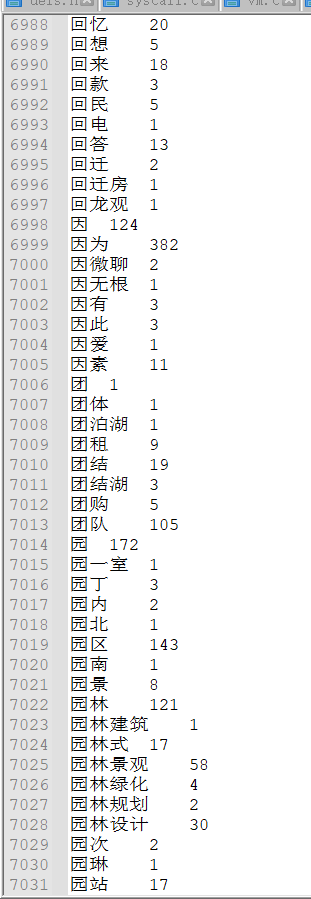


# 实验结果与分析

通过实验，以小组的三台机器完成了hadoop分布式环境的搭建，其中master节点即使namenode节点也是datanode节点，slave1是datanode节点包含了resourcemanager功能，slave2是secondary namenode节点和datanode节点.利用hadoop的分布式存储系统hdfs存储了了爬虫爬取并分词后的结果文件，最终用分布式计算mapreduce功能完成了词频的统计。



搭建成功截图



个人签名 ： 王铭

2021年 11 月 26 日